Piecing together the biogeographic history of *Chenopodium vulvaria* L. using botanical literature and collections

Quentin J Groom

This study demonstrates the value of legacy literature and historic collections as a source of data on environmental history. *Chenopodium vulvaria* L. has declined in northern Europe and is of conservation concern in several countries, whereas in other countries outside Europe it has naturalised and is considered an alien weed. In its European range it is considered native in the south, but the northern boundary of its native range is unknown. It is hypothesised that much of its former distribution in northern Europe was the result of repeated introductions from southern Europe and that its decline in northern Europe is the result of habitat change and a reduction in the number of propagules imported to the north. An historical analysis of its ecology and distribution was conducted by mining legacy literature and historical botanical collections. Text analysis of habitat descriptions written on specimens and published in botanical literature covering a period of more than 200 years indicate that the habitat and introduction pathways of *C. vulvaria* have changed with time. Using the non-European naturalised range in a climate niche model it is possible to project the range in Europe. By comparing this predicted model with a similar model created from all observations it is clear that there is a large discrepancy between the realized and predicted distributions. This is discussed together with the social, technological and economic changes that have occurred in northern Europe, with respect to their influence on *C. vulvaria*.

1   Author: Quentin J. Groom

2   Affiliation: Botanic Garden Meise, Bouchout Domain, Nieuwelaan 38, 1860 Meise, Belgium

3   Email: quentin.groom@br.fgov.be

Introduction

Legacy biodiversity literature is a potential source of useful information on the past distributions of organisms. While these texts have always been available in academic libraries, their accessibility and discoverability has been significantly enhanced by projects such as the Biodiversity Heritage Library (www.biodiversitylibrary.org) and other online digital sources. The ability to search a whole corpus of historical literature for a Latin name of an organism dramatically increases the accessibility of this information and makes literature searches possible that once would have been unfeasible. In parallel, the widespread digital imaging of herbarium specimens and transcription of their labels has also made these data considerably more accessible, which, combined with historic literature, has created a large pool of information from which the phytogeographic historian can draw evidence (Vellend et al., 2013).

Chenopodium vulvaria L., is a small, inconspicuous species that grows largely in places disturbed by humankind. It is not remarkable morphologically, but it is nonetheless distinctive due to its foul smell, which is described as similar to that of rotten fish. Its distinctiveness makes it particularly suited to a study using historic literature, because there is less concern that published accounts refer to other species as a result of misidentification.

C. vulvaria is currently a red-listed species in several countries including Sweden (www.artfakta.se), the United Kingdom (Cheffings et al., 2005), Belgium (Kestemont, 2010), Luxembourg (Colling, 2005), Czech Republic (Grulich, 2012) and some regions of France (Ferrez, 2005). In Great Britain, Ireland and Flanders, comparisons of atlas data show that *C. vulvaria* is in severe decline (Preston et al., 2002; Van Landuyt et al., 2006). In contrast, it has naturalised in California (Calflora, 2014), Argentina (Planchuelo, 1975; Giusti, 1997), Chile (Boelcke et al., 1985) and Australia (Atlas of Living Australia, 2014).

C. vulvaria is widespread in countries bordering the Mediterranean and eastward to Afghanistan and Mongolia (Jalas & Suominen 1980; Meusel, Jäger & Weinert, 1992). Yet it is clear from historical literature and specimens that it was common in parts of northern Europe during the 18th and 19th centuries. Turner (1548) wrote "*It groweth muche aboute the walles in Bon in Germany*"; Bucher (1806) wrote in the Flora of Dresden "*An den strassen der vorstadt und sonst gemein*" translated as "By the streets of suburbs and usually common"; Curtis (1777) stated "*This species is very common in the neighbourhood of London…*" and Hooker (1821), in his flora of Scotland, describes it as "*frequent*". Lejeune and Courtois (1828) described its distribution in Belgium as

35   "*…per totum Belgium passim*", which can be translated as "everywhere throughout Belgium" and

36   in France it was described as "*commun*" and "*trés commune*" (i.e. "common" and "very

37   common") on specimens from Narbonne (1846, P04922786) and Paris (1819, P05292341) in the

38   Museum National d'Histoire Naturelle.

39   The native distribution of *C. vulvaria* is unknown and its long association with anthropogenic

40   disturbance makes this impossible to determine. Floras in Northern and Central Europe variously

41   describe it either as a native or an archaeophyte, though the evidence for categorizing it in either

42   category is slim and is probably based on the anthropogenic habitats that *C. vulvaria* often

43   inhabits.

44   Many other members of the Amaranthaceae live in disturbed, nutrient rich habitats and may be

45   halophytic. *C. vulvaria* itself is often found in disturbed, eutrophic and coastal habitats. In

46   general, species of such habitats are increasing and spreading in northern Europe (Wróbel,

47   Tomaszewicz & Chudecka, 2006; Van Landuyt et al., 2008; Smart et al., 2003; Šerá, 2011;

48   Groom, 2013). So at face value, *C. vulvaria* appears well adapted to modern habitats in Europe

49   and yet it has declined.

50   One possible explanation for its apparent decline in northern and central Europe may be a

51   misunderstanding of its former occurrence, its presence in the north being the result of propagule

52   pressure from its heartland in southern Europe, constantly reinforcing the introduced populations

53   in the north. One or many introduction pathways may have existed that delivered *C. vulvaria* seed

54   outside of its normal range and these pathways have since reduced in importance, causing a

55   collapse in the population. Another possible explanation is change to its former habitat, though

56   the details of its ecology are too poorly known to understand what these changes may have been.

57   For non-woody plants there are few sources of data to examine recent biogeographic change.

58   Palynology and the study of archaeological remains can be useful, but many species do not have

59   sufficiently distinctive characters to identify them from their remains. In these cases, historical

60   literature and collections may be the only sources of data on their former habitats and locations.

61   Given the shortage of data, an alternative approach, widely used to model the potential

62   distribution of organisms, is bioclimatic modelling. Many studies have used observations from

63   the known native range of a species to extrapolate its potential invasive range (e.g. Macfadyen &

64   Kriticos, 2012). In ecological theory the potential bioclimatic range is generally considered to be

65   larger than the realized distribution as a consequence of additional non-climatic limitations to

66   distribution, such as edaphic factors (Araujo & Peterson, 2012). However, in the case of *C.*

67   *vulvaria* the native range is not known and frequent non-persistent introductions mean that the

68   realized distribution predicted from observations may be larger than its true bioclimatic range.

69   For *C. vulvaria* the location of naturalisation in Australia, North America and South America

70   might be a clearer indication of its bioclimatic range than within Europe, where it is hard to

71   distinguish established from casual occurrences. Assuming that this species is well established

72   and stable in its alien range, we can use the known naturalised range to model the climate

73   envelope and extrapolate this to Europe to identify the areas where the climate is suitable for *C.*

74   *vulvaria*. In this manner we can indicate those places where this species has been observed but is

75   unlikely to be persistent. A similar approach has been used by Zhu et al. (2012) to predict the

76   source of the invasive insect *Halyomorpha halys* (Stål, 1855) and has been used by others to

77   model the historical distributions of trees (Benito-Garzón et al., 2008). This approach assumes

78   niche conservatism, but such conservatism is apparently common place and perhaps typical of

79   invasive terrestrial plant species (Townsend Peterson, 2011; Petitpierre et al., 2012)

80   My hypothesis is that *C. vulvaria* was formally more abundant in northern Europe and its current

81   decline is the result of changes in the introduction pathways and loss of habitat. In this study I

82   draw on botanical literature and specimens to identify habitat change and historic introduction

83   pathways. I use text analysis of habitat descriptions to demonstrate how its habitat has changed

84   over the past 200 years and I use bioclimatic niche modelling to contrast the realized range within

85   Europe with the projected range based upon naturalised occurrences outside Europe.

86   **Methods**

87   Observation and specimen details were collected in a Common Data Model (CDM) database

88   which is the central component of the EDIT Platform for Cybertaxonomy (Ciardelli et al., 2009;

89   Berendsohn et al., 2011). Two methods were used to extract observations from literature, either

90   XML markup or direct data entry. Digitised treatments were marked up with XML using the

91   GoldenGate editor (http://plazi.org/?q=GoldenGATE, Sautter, Böhm & Agosti, 2007); uploaded

92   to the PLAZI taxonomic treatment repository (plazi.org) and imported to the CDM database.

93   Alternatively the observation details were copied from the treatment and entered manually into

94   the CDM database using the EDIT Taxonomic Editor (Ciardelli et al., 2009). Observations were

95   collected from the biodiversity literature by reading the Biodiversity Heritage Library corpus

96   systematically after searching for *C. vulvaria* L. and its synonyms *C. foetidum* Lam., *C. olidum*

97   Curt., *Atriplex vulvaria* Crantz and *Vulvaria vulgaris* Bubani. Other published observations were

98   gathered from publications in the Library of the Botanic Garden, Meise. A list of the sources of

99   observations of *C. vulvaria* is available in supplementary file S1. A complete survey of non-

100   digitised literature is unfeasible, but there was an effort to check multiple floras of every

101   European country and any other country with a temperate climate suitable for *C. vulvaria*.

102   Digitised observation data were also gathered from databases, primarily from the Global

103   Biodiversity Information Facility (GBIF) (data.gbif.org, accessed 08 Nov 2013; see

104   supplementary file S2), but also from the Atlas of Living Australia (2014); the Botanical Society

105   of Britain and Ireland (2013) and Herbaria@home (2013). Scientific articles and websites

106   containing observations were also discovered using search engines (scholar.google.be;

107   google.be). Data were standardised and imported directly into the CDM database.

108   Specimen data were gathered from herbaria by transcription of label information. Specimens

109   from 31 herbaria are included in the study; their abbreviations follow the Index Herbariorum

110   (http://sciweb.nybg.org/science2/IndexHerbariorum.asp): ABS, ALA, B, BBB, BC, BIRM, BM,

111   BR, C, CONC, FABR, FRU, HFLA, K, L, LISU, M, MANCH, MW, NYS, P, PRC, RNG, SO,

112   SLBI, SOM, SOMF, UBC, UC, WAG, WU and others contributing data to GBIF (Herbaria

113   names in supplementary file S2). Many other herbaria and herbarium catalogues were searched

114   without finding specimens and several herbaria were contacted and either contained no

115   specimens or did not respond. Undoubtedly there are more specimens and observations of *C.*

116   *vulvaria* to be discovered, but I believe these to be a representative sample and a large proportion

117   of those that exist. Undated specimens were not used in the study; however, it is usual for

118   published observations to be undated and therefore the publication date was used instead. In

119   studying biographical information of collectors it is clear that most undated observations in old

120   floras are within 35 years of the publication date and authors tend to provide dates when those

121   dates are a long time before the publication date. In total 2456 observations were collected from

122   specimens and literature. These data span 465 years from 1548 to 2013, though there are only

123   two observations from the 16th century, two from the 17th century and nineteen from the 18th

124   century.

125   **Text analysis of habitats**

126   The text describing the habitat of *C. vulvaria* was collected from 104 floras, 33 scientific articles,

127   119 specimens and 5 websites, covering the years 1787 to 2014. The texts were written in 12

128   languages, English (35%) German (20%), French (17%), Latin (12%), Dutch (4%), Italian (3%),

129   Portuguese (3%), Spanish (3%), Hungarian (1%), Danish (1%), Catalan (1%) and Czech (<1%).

130   Each description was broken down into tokens consisting of either single words or short phrases

131   describing a single aspect of the habitat. Thus the description "*In Straßen, an Häusern,*

132   *Stallungen, Düngerstätten*" was broken down into the tokens "*Straßen*" (roads), "*an Häusern*"

133   (near houses), "*Stallungen*" (stables) and "*Düngerstätten*" (manure heaps). This process created

134   475 habitat tokens. These tokens were then translated to English using native speakers of English,

135   German, French and Dutch and for other languages a combination of Google Translate

136   (translate.google.be) and the multilingual collaborative dictionary Wiktionary (wiktionary.org).

137   To conduct the analysis it was necessary to reduce the number of habitat terms, which was done

138   in two stages. The anglicized tokens were first simplified to closely related terms. Thus the terms

139   "*by foot of the city walls*", "*along walls*", "*under walls*", "*mud walls*", "*foot of walls*", "*under*

140   *walls*", "*under a wall*" and "*foot of the church yard wall*" were all replaced by "*by walls*". This

141   process reduced the number of habitat words to fifty. These fifty words were then arranged into

142   logically related categories. Thus "*by walls*", "*by fences*" and "*by hedges*" were grouped together

143   under the term "*boundaries*". This reduced the number of habitat categories to fifteen: animal

144   waste, boundaries (including walls), coastal, disturbed and grazed land, dry & bare soil,

145   habitation, hills, horticulture, industry, rail, roads, sand and rock, shipping, waste, wetland. A full

146   list of the tokens contributing to each category is provided in the supplementary material (Table

147   S3). Throughout the process the tokens were kept associated with the date; either the year the

148   specimen was collected, observed or the year of publication. To analyse the use of habitat words

149   in the collected corpus, the simplified habitat terms were pooled into 20 year periods from 1780

150   onwards. The proportional use of each habitat term was then calculated for each period.

151   Statistical analysis was conducted in R (version 2.15.2) using generalized linear modelling with

152   binomial errors, weighted with the number of tokens contributing to each pool. All models were

153   checked for overdispersion using the ratio of the residual deviance and the degrees of freedom,

154   but none were found to be overdispersed.

155   **Analysis of distribution**

156   Except for the rare occasions when coordinates where available with the specimen or

157   observation, georeferencing was carried out manually according to best practise (Chapman &

158   Wieczorek, 2006). Error radii for coordinates were not available for most records in databases,

159   but they were estimated for the coordinates georeferenced in this study, however, they were not

160   used to select data for the analysis. The average error radius was 11 km and the mode and median

161   were both 10 km. *C. vulvaria* is a largely lowland species and errors in georeferencing of these

162   magnitudes are insignificant for bioclimatic modelling at a global scale compared to the other

163   inherent biases in these data.

164   Species distribution modelling was conducted using the BioVel Ecological niche modelling

165   workflow and services (www.biovel.eu). The ecological niche modelling workflows were run on

166   6th Aug 2014. The workflow uses the Maxent method based upon Phillips, Dudík & Schapire

167   (2004) and using the openModeller web service (de Souza Muñoz et al., 2011). Models were

168   created using the default parameter and all 19 layers of the WorldClim global climate layers 10

169   arc minutes, version 1.4, release 3 (Hijmans et al., 2005).

170   Non-European observations used for modelling were only those locations where it was clear,

171   either from the notes on the specimens or from Floras, that the species forms persistent

172   population at these sites and was alien. If there was any doubt to the status, modern Floras were

173   consulted to ascertain the persistence of the species in the area. Seventy observations were

174   collected from locations outside Europe in southern Argentina; California; Chile; New Zealand;

175   South Australia; South Africa; Tasmania; Tierra del Fuego (Argentina) and Victoria (Australia).

176   The status of *C. vulvaria* in South Africa and New Zealand is not clear and therefore the data was

177   not included. It is also believed to be native to Mongolia but only one observation was found.

178   After selection, 42 observations from the naturalised range were used to model the range,

179   however, weeding of duplicates during the workflow reduced the number to 32. The dates of

180   these records were from 1863 to 2012, although 86% dated from 1950 onward. To model the

181   realized range, all global observations were used, which resulted in 1894 observations after

182   weeding of duplicates.

183   ## Results

184   ### Text analysis

185   Four habitat categories were notably more frequent than the others (Fig. 1). These categories are

186   firstly waste, including rubbish piles, rubble, ruins and waste places of all kinds; secondly,

187   boundaries, mainly at the base of walls; thirdly, roads and roadsides, including streets and farm

188 tracks; and fourthly horticulture, such as gardens and other cultivated places. The habitat

189 categories in Fig. 1 are not mutually exclusive, but often describe different aspects of the same

190 habitat such as the proximity to landscape features, soil type, nutrient status and moisture.

191 In summary, the habitat analysis underscores several aspects. *C. vulvaria* is strongly associated

192 with humankind; natural habitats such as coastal habitat and wetlands are mentioned infrequently.

193 *C. vulvaria* is intolerant of competition; none of its habitats are defined by established vegetation

194 types, such as meadows, woodland or heaths. It is frequently associated with transport routes and

195 it is usually associated with some form of soil disturbance.

196 When the use of these terms was compared over time, no significant change was found for the

197 use of terms relating to animal waste, coastal, dry & bare soil, habitation, hills, horticulture,

198 industry, rail, roads, shipping and waste. Figure 2 shows the changes of eight of these categories,

199 including the only four where there were significant changes. The significant changes were

200 increases in the proportion of the terms related to wetland ($p < 0.01$, DF=11), sand and rock ($p <$

201 0.05, DF=11) and disturbed and grazed land ($p < 0.05$, DF=11), whereas there has been a

202 significant decrease in the proportion of terms related to boundaries ($p < 0.001$, DF=11). Of these

203 significant changes only terms relating to boundaries were also highly frequent in the corpus (Fig.

204 1).

205 Those 194 habitat tokens that also had a geolocation were plotted on a map to show the spread of

206 habitat categories (Fig. S4). Habitat categories were widely distributed rather than being clustered

207 in single counties or in areas of a common language. There is no indication of bias in the

208 translation and the categorization of habitats. It also implies that the results of the habitat analysis

209 are broadly applicable.

210 **Introduction vectors, pathways and origins**

211 Not surprisingly, clear expressions of the introduction vector were rare on specimens and in

212 publications. Where introduction vectors were evident they are summarised in Table 1. Ballast

213 soil at ports was the earliest vector mentioned in the corpus and it was also most frequently

214 mentioned. However, it stops being mentioned in the early 20th century. Several specimens and

215 observations implicate the transport of ore. *C. vulvaria* was reported on Chromite in Baltimore,

216 USA between 1953–1958 (Reed, 1964); in Norway in 1954 (Uotila, 2001); on manganese ore in

217 Norway between 1931–1935 and near an ore crushing plant in Kyrgyzstan in 1961 (Lazkov,

218 Sennikov & Naumenko, 2014). Various agricultural products are mentioned as possible vectors of

219 introduction, such as grain. However, there is no mention of it being introduced with other
220 produce commonly imported from the Mediterranean such as herbal medicines or tobacco, even
221 though *C. vulvaria* is frequently associated with waste.

222 Evidence for the pathways of introductions is scant, but shipping and railways are mentioned.
223 Although roads are the most frequently mentioned transport system (Fig. 1), it is unclear if the
224 presence of this species on roads relates to the introduction pathway or whether roads just provide
225 suitable habitat.
226 Evidence for the origin of introductions is also slim, although where the origin is mentioned it is
227 always from a country in the Mediterranean region (Uotila, 2001). There is no evidence of return
228 introductions from naturalised populations outside Europe.

229 ### Comparing the fundamental and realized climatic niche
230 The observations of *C. vulvaria* within Europe are from an inseparable mixture of stable
231 populations and casual occurrences. It is therefore impossible to validate a model for the
232 fundamental climatic niche of *C. vulvaria*. For this reason I did not attempt to refine the output of
233 the models by adjusting their default parameters or by eliminating climate layers. It is
234 nevertheless informative to contrast models created from the known naturalised range outside
235 Europe with the realized range within Europe (Fig. 3, Fig. 4). The actual climatic niche, predicted
236 from observations from the naturalised range outside Europe predicts the presence of *C. vulvaria*
237 in southern and western Europe, North Africa and the Middle-east, notably, Spain, western
238 France and Turkey (Fig. 3). The actual observations and the climate niche model created from
239 them show a much wider distribution, which extends much further north and eastward than the
240 niche model created from the naturalised range (Fig. 3, Fig. 4). Figure 5 shows the locations of
241 actual observations and the dates they were made. It demonstrates that there has been a general
242 decline in the number of observations from northern Europe, but it also suggests unevenness in
243 surveying effort between different countries and different time periods.

244 ## Discussion
245 This study tracks the distribution and habitat changes of *C. vulvaria* over more than 200 years.
246 Over this period botanical literature becomes more common and sufficiently abundant for
247 analysis. Simultaneously, botanical specimens became more frequently collected and better
248 documented, further adding to the analysable corpus of historical documents.

249   Over the past two centuries many social, economic and technological changes have occurred that

250   may have influenced the abundance and distribution of *C. vulvaria*. Some key events in this

251   period are the expansion of the railway network in the 19th century, the adoption of motorised

252   road transport in the early 20th century, the decline in the uses of horses for transport and

253   agriculture in the 20th century; the transition from sail to steam powered ships at the turn of the

254   20th century; the discovery of herbicides in the mid-20th century and the Green Revolution in the

255   latter half of the 20th century. *C. vulvaria* to some extent benefits from its association with

256   humankind, however, for the same reason it will be more acutely affected by cultural and

257   technological changes than many other species.

258   Key habitat features of *C. vulvaria* were identified though text analysis. This species has been,

259   and still is, strongly associated with humankind, both as a weed of cultivation and as a ruderal

260   plant. The analysis identifies habitat traits such as its avoidance of competition and the

261   association with waste. The genus *Chenopodium* is considered to be nitrophilous (Ellenberg et al.,

262   1991); indeed *C. vulvaria* is sometimes associated with habitats linked to animal dung, however,

263   it is much more commonly associated with other types of waste or cultivated place (Fig. 1).

264   The temporal analysis of habitat change indicates that *C. vulvaria* is still associated with many of

265   the same habitats as it was in the past, such as agriculture, transport and waste (Fig. 2). However,

266   in the 20th century habitat descriptions have included proportionally more words related to natural

267   or semi-natural habitats, such as grazing, sand and wetland.

268   The reference to wetland amongst the habitats needs further explanation, because *C. vulvaria* is

269   not a typical wetland plant. It does not grow in water, but colonises bare soil exposed in the

270   summer at the margins of rivers, ditches and lakes. Thus its association with wetland is of an

271   opportunistic colonizer of habitats free from competition, rather than a true wetland plant.

272   The habitat where *C. vulvaria* has declined is along boundaries, particularly along walls, which

273   contributed 80% of the boundary terms. *C. vulvaria* does not grow on walls, but beside them,

274   which appears at first sight to be a rather non-specific habitat description. However, the margins

275   of walls have changed considerably in the past 200 years. Walls were once built using lime

276   mortar, rather than cement, and were frequently painted with whitewash, a mixture of calcium

277   hydroxide and chalk. Whitewash gave the traditional white or pink colour to houses throughout

278   Europe. Consequently, the soil in the immediate vicinity of walls would have been alkaline. *C.*

279   *vulvaria* is not known as an alkaliphile, however it is clearly tolerant of high pH as it has been

280   found on the ultrabasic rock chromite (Reed, 1964). Furthermore, because horses were used for

281 transport and farm animals were driven along roads, the base of walls would have been strewn

282 with animal waste. Such fertile alkaline habits do not occur by walls in modern towns and it is

283 speculated that the technical changes in building practises and changes to transportation have

284 contributed to the decline of *C. vulvaria*.

285 Text analysis is clearly a useful tool for environmental historians, nevertheless, it is susceptible to

286 the fallibility and biases of authors, who may uncritically follow their forbears or write from

287 hearsay rather than experience. Also, botanical activity is spatially and temporally biased

288 (Delisle et al., 2003; Schmidt-Lebuhn et al., 2013). For example, British and German botanical

289 literature has, and continues to be, more abundant than for other countries in Europe.

290 Compared to the analysis of habitat, evidence for introduction vectors, pathways and origins was

291 limited. The results, however, suggest that there were multiple vectors introducing *C. vulvaria* to

292 northern Europe, but particularly as a grain contaminant and in ship's ballast. The frequent

293 occurrence of *C. vulvaria* in waste perhaps indicates that its seeds were contaminants of many

294 crops. Indeed, different descriptions accompanying specimens mentioned *C. vulvaria* in crops of

295 lentils and potatoes. Unfortunately, the source of a casual introduction is rarely obvious by the

296 time the plant is mature. Weed species that are dispersed as seed contaminants have declined

297 throughout Europe in the 20th century; this is, in part, a consequence of improved seed cleaning

298 methods (Hilbig, 1987; Sutcliffe & Kay, 2000; Lososová, 2003). Most of these weeds are

299 considered archaeophytes to northern Europe, typical examples are *Agrostemma githago* L.

300 *Glebionis segetum* (L.) Fourr. and *Lithospermum arvense* L.

301 Soil was used as ballast on sailing ships during the 18th and 19th centuries to provide stability to

302 cargo ships when not carrying heavy loads. In ports, where heavy materials were loaded, ballast

303 was removed and replaced by cargo. Large hills of ballast soil were a common feature of busy

304 ports, particularly in areas of mining and heavy industry, such as in northern Europe. These

305 ballast hills were a large reservoir of propagules for many species (Carlton, 2011). The large

306 number of specimens and observations reflects the importance of this invasion pathway, but

307 might be somewhat over-represented because botanists were attracted to ballast heaps as a source

308 of novel species and because the vector of the propagules is clear in this case.

309 Ore is also mentioned as an introduction vector to the USA and Norway. Chromium processing

310 began in Baltimore, USA in 1822, at which time only local chromium ore deposits were

311 processed (Newcomb, 1994). However, by the end of the 19th century local chromium deposits

312   were exhausted and processing continued with imported ore until the end of the 20ᵗʰ century.

313   Similarly, Norway is also a large processor of imported chromium ore, for example in 1992 the

314   country imported 187,965 tonnes of chromite ore from Turkey (Plachy, 1992). Indeed, it is likely

315   that some of the chromite imported into Baltimore was also from Turkey where chromite was

316   first mined in the 19ᵗʰ century (Zengin, 1957). Therefore, it seems that exports of chromite from

317   Turkey could have been a pathway for dispersal of *C. vulvaria* during the 20ᵗʰ century.

318   Animal dung is often mentioned as a growing medium for *C. vulvaria*, which is indirect evidence

319   for endozoochory. Certainly, other *Chenopodium* species are dispersed in this manner and *C.*

320   *vulvaria* is eaten by ruminants despite its smell (Withering 1776; Haarmeyer et al., 2010). In the

321   21ˢᵗ century yet another vector of *C. vulvaria* introduction has been created, that of imported olive

322   trees (Hoste et al., 2009). These mature trees are extracted from olive groves with a large amount

323   of soil and are sold in northern Europe as horticultural novelties.

324   Though dispersal vectors are rarely mentioned in the corpus, it is clear that *C. vulvaria* has been

325   dispersed by a wide variety of vectors and through a number of pathways (Table 1). There are

326   historic periods associated with each vector and if this analysis was extended to more species,

327   one would be able to further refine the time frames during which these pathways were operating.

328   From the diversity of distribution vectors it is clear that *C. vulvaria* has been widely introduced

329   outside its natural climatic range and it often grows temporarily. However, with the exception of

330   horticultural imports, introduction pathways of *C. vulvaria* declined midway through the 20ᵗʰ

331   century.

332   The sporadic occurrence of *C. vulvaria* presents a problem for the selection of occurrences for

333   distribution modelling. Unless all casual occurrences are eliminated from the data before fitting,

334   the model would indicate a much broader climatic range. Separating permanent populations from

335   casual occurrences is impossible for Europe where anthropogenic disturbance and trade have

336   confused the quasi-natural distribution. However, in the naturalised range the situation is much

337   clearer. Most, if not all, modern observations of *C. vulvaria* in California, Australia and South

338   America appear to be from naturalised populations, that is to say, the associated meta-data

339   indicates the presence of a population and there is no indication of a recent introduction.

340   Therefore, the naturalised distribution outside Europe should reflect the fundamental climatic

341   niche of the species, as long as the distribution is at equilibrium. This assumption seems

342   reasonable since old casual records of *C. vulvaria* occur throughout the world, but naturalised

343   populations persist in only a few of those places. Clearly, introduction events were occurring all

344 over the world for several hundred years of international trade, but *C. vulvaria* only naturalised in

345 a few of those places where habitat and climate were suitable.

346 Projecting the bioclimatic range in Europe from naturalised alien populations elsewhere predict a

347 more south-western distribution of *C. vulvaria*, than the modelling using all occurrences (Fig. 3,

348 Fig. 4). Yet, these rather crude models indicate that the naturalised distribution of *C. vulvaria* has

349 a climate much closer to that of southern Europe and North Africa than to northern and central

350 Europe. The distribution models are consistent with my hypothesis that historically *C. vulvaria*

351 was only present in parts of northern Europe because of repeated introductions, and that, in these

352 places, the climate is unsuitable for lasting populations to exist.

353 This conclusion assumes niche conservatism between the native and introduced populations. It

354 further assumes that the whole of the fundamental niche climate space is available to *C. vulvaria*

355 in its introduced range (Soberón and Peterson, 2011). Therefore, an alternative hypothesis is that

356 *C. vulvaria* has undergone a climatic niche shift in its introduced range such as suggested by

357 Gallagher et al. (2010). However, given that this would have had to occur in parallel on three

358 continents, this seems less plausible.

359 Discrepancies between the projected model and the realized distribution could be the result of

360 several factors, either an incorrect model; lack of suitable habitat; spatial variation in surveying

361 effort, or plants growing outside their actual climatic niche due to local factors. The model

362 projecting distribution from the naturalised range is based on relatively few observations and

363 would be improved by more data (Feeley & Silman, 2011). Nevertheless, any distribution model

364 of this species has to address the problem of casual occurrences. The shortage of observations

365 from countries such as Turkey and Morocco, in apparent contrast to the models, is at least in part

366 due to lack of collecting in these regions, but also due to the inaccessibility of the data from these

367 countries. These results are a good reminder to those who would extrapolate native ranges onto

368 potentially invasive ranges that it is not always possible to predict the naturalised distribution of

369 an invasive species based upon fundamental climatic niche both because of the lack of data and

370 indistinct native range boundaries, but also because climate is not the only determinate of

371 distribution.

372 **Conclusions**

373 Text analysis is a useful technique to study recent ecological and distributional change. Despite

374 its limitations it provided information, which would be difficult if not impossible to obtain from

375 other sources. As a larger volume of semantically enhanced biodiversity literature becomes

376 available it will allow much more sophisticated habitat analysis covering many more species. The

377 ability to contrast data from different species will strengthen results and allow correction for

378 some of the biases. Furthermore, the development of environmental ontologies and thesauri will

379 simplify the method and improve repeatability (Buttigieg et al., 2013). This will allow over-

380 representation analysis of ontological terms associated with one species compared to the

381 incidence of these terms in the whole corpus.

382 Analysis of environmental descriptions indicates that the habitat of *C. vulvaria* has changed over

383 the past two centuries, particularly next to walls. Multiple vectors and pathways have been

384 involved in the human mediated dispersal of *C. vulvaria*, but different vectors and pathways were

385 active in different periods. In the past *C. vulvaria* would have been dispersed to many places

386 outside of its climatic niche. It is reasonable to believe that many of the observations of *C.*

387 *vulvaria* in northern Europe were the result of introductions and that a reduction in the propagule

388 pressure in recent years has consequently lead to a decline in observations of this species. It is

389 concluded that humankind spread *C. vulvaria* to northern Europe and created habitat for it to

390 grow and then inadvertently removed the habitat and the introduction pathways causing a decline.

Acknowledgements

392 I would like to thanks the following people for their help accessing collections and transcribing

393 specimen details: Ana Isabel D. Correia of the Faculdade de Ciências da Universidade de Lisboa;

394 Prof. Dr. Roberto Rodríguez Ríos from the Universidad de Concepción, Chile; Noemí Montes

395 and Neus Ibáñez Cortina of the Institut Botànic de Barcelona; Sabrina Eckert, Anton Güntsch,

396 Patricia Kelbert, Andreas Müller and Robert Vogt of the Botanischer Garten und Botanisches

397 Museum Berlin-Dahlem; Vladimir Vladimirov from the Bulgarian Academy of Sciences; Michal

398 Štefánek of Charles University in Prague; Alan Paton from the Royal Botanic Gardens, Kew;

399 Leni Duistermaat & Dr. Hubert Turner of Naturalis Biodiversity Center and the specimen

400 digitization team at the Botanic Garden Meise and Sabine Metzger. This work has been supported

401 by the pro-iBiosphere project, which have been funded by the European Commission under the

402 7th Framework Programme (grant agreement number 312848).

References

404 Araujo MB, Peterson AT. 2012. Uses and misuses of bioclimatic envelope modeling. *Ecology*

405 93:1527–1539.

406 Atlas of Living Australia. 2014. *Available at*

407   http://bie.ala.au/species/Chenopodium+vulvaria (accessed 25 February 2013).

408   Benito-Garzón M, Sánchez de Dios R, Sáinz Ollero H. 2008. The evolution of the *Pinus*

409   *sylvestris* L. area in the Iberian Peninsula from the last glacial maximum to 2100 under climate

410   change. *The Holocene* 18:705–714.

411   Berendsohn WG, Güntsch A, Hoffmann N, Kohlbecker A, Luther K, Müller A. 2011.

412   Biodiversity information platforms: From standards to interoperability. *ZooKeys* 150:71–87.

413   Boelcke O, Correa NM, Moore DM, Roig FA. 1985. Catálogo de las Plantas Vasculares. In:

414   Boelcke O, Moore DM, Roig FA, eds., *Transecta Botánica de la Patagonia Austral. Proyecto*

415   *Internacional*. Buenos Aires: Consejo Nacional de Investigaciones Científicas y Técnicas, 129–

416   255.

417   Botanical Society of Britain and Ireland. 2013. Distributions Database. *Available at*

418   http://bsbidb.org.uk. (accessed 21 February 2013).

419   Bucher, C.T. (1806) *Florae Dresdensis nomenclátor*. Dresden: Walther.

420   Burk I. 1877. List of plants recently collected on ships' ballast in the neighborhood of

421   Philadelphia. *Proceedings of the Academy of Natural Sciences of Philadelphia* 29:105–109.

422   Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE, and the ENVO Consortium. 2013.

423   The environment ontology: contextualising biological and biomedical entities. *Journal of*

424   *Biomedical Semantics* 4:43.

425   Calflora. 2014. Information on California plants for education, research and conservation, based

426   on data contributed by dozens of public and private institutions and individuals, including the

427   Consortium of Calif. Herbaria. Berkeley (USA). *Available at* www.calflora.org (accessed 3

428   March 2014).

429   Carlton JT. 2011. Ballast. In: Simberloff D, Rejmanek M, eds. *Encyclopedia of Biological*

430   *Invasions*. Berkeley and Los Angeles: University of California Press, 43–49.

431   Chapman AD, Wieczorek J. 2006. Guide to best practices for georeferencing. Copenhagen:

432   Global Biodiversity Information Facility, 1–77.

433   Cheffings CM, Farrell L, Dines TD, Jones RA, Leach SJ, McKean DR, Pearman DA, Preston

434   CD, Rumsey FJ, Taylor I. 2005. The vascular plant red data list for Great Britain. *Species Status*

435   7:1–116.

436   Ciardelli P, Kelbert P, Kohlbecker A, Hoffmann N, Güntsch A, Berendsohn WG. 2009. The EDIT

437   Platform for Cybertaxonomy and the Taxonomic Workflow: Selected Components. In: Fischer S,

438   Maehle E, Reischuk R, eds. *INFORMATIK 2009, Im Focus das Leben, Beiträge der 39.*

439   *Jahrestagung der Gesellschaft für Informatik e.V. (GI), 28.9. - 2.10. in Lübeck*. Lecture Notes in

440   Informatics (LNI) 154 - S. 28: 625–638.

441   Colling G. 2005. Red list of the vascular plants of Luxembourg. *Ferrantia* 42:80.

442   Curtis W. 1777. *Flora Londinensis vol. 5*. London: William Curtis.

443   de Souza Muñoz M, De Giovanni R, de Siqueira M, Sutton T, Brewer P, Pereira R, Canhos DAL,

444   Canhos VP. 2011. openModeller: A generic approach to species' potential distribution modelling.

445   *GeoInformatica* 15:111–135.

446   Ellenberg H, Weber HE, Düll R, Wirth V, Werner W, Paulissen D. 1991. Zeigerwerte von

447   Pflanzen in Mitteleuropa. *Scripta Geobotanica* 18:1–248.

448   Feeley KJ, Silman MR. 2011. Keep collecting: accurate species distribution modelling requires

449   more collections than previously thought. *Diversity and Distributions* 17:1132–1140.

450   Ferrez Y. 2005. Liste rouge de la flore vasculaire menacée ou rare de Franche-Comté Proposition.

451   *Les nouvelles archives de la flore jurassienne* 3:217–229.

452   Gallagher RV, Beaumont LJ, Hughes L, Leishman MR. 2010. Evidence for climatic niche and

453   biome shifts between native and novel ranges in plant species introduced to Australia. *Journal of*

454   *Ecology* 98:790–799.

455   Groom QJ. 2013. Some poleward movement of British native vascular plants is occurring, but the

456   fingerprint of climate change is not evident. PeerJ 1:e77.

457   Grulich V. 2012. Red list of vascular plants of the Czech Republic: 3rd edition. *Preslia* 84:631–

458   645.

459   Giusti L. 1997. Chenopodiaceae. In: Hunziker AT, ed. *Flora fanerogámica argentina.* Córdoba:

460   Proflora-Conicet: 40: 1–53.

461   Haarmeyer DH, Bösing BM, Schmiedel U, Dengler J. 2010. The role of domestic herbivores in

462   endozoochorous plant dispersal in the arid Knersvlakte, South Africa. *South African Journal of*

463   *Botany* 76:359–364.

464   Herbaria@home. 2013. *Available at http://herbariaunited.org/atHome/* (accessed 21 February

465   2013).

466   Hilbig W. 1987. Changes in segetal vegetation under conditions of industrialized agriculture.

467   *Archives of Nature Conservation & Landscape Research* 27:229–249.

468   Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated

469   climate surfaces for global land areas. *International Journal of Climatology* 25:1965–1978.

470   Hjelt H. 1906. Caryophyllaceae-Resedaceae. In: *Conspectus Florae Fennicae in Acta Societatis*

471   *pro Fauna et Flora Fennica 30*. Finland: Helsingforsiae.

472   Hooker WJ. 1821. *Flora Scotica*. London: Archibald Constable and Co. & Hurst, Robinson, and

473   Co.

474   Hoste I, Verloove F. Nagels C, Andriessen L, Lambinon J. 2009. De adventievenflora van in

475   België ingevoerde mediterrane containerplanten. *Dumortiera* 97:1–16.

476   Jalas J, Suominen J. 1980. *Atlas Florae Europaeae, Vol. 5*. Helsinki: The Committee for Mapping

477   the Flora of Europe and Societas Biologica Fennica Vanamo.

478   Kestemont B. 2010. A red list of Belgian threatened species. Brussels: Statistics Belgium.

479   Delisle F, Lavoie C, Martin J, Lachance D. 2003. Reconstructing the spread of invasive plants:

480   taking into consideration biases associated with herbarium specimens. *Journal of Biogeography*

481   30:1033–1042.

482   Lazkov G, Sennikov A, Naumenko A. 2014. New records in vascular plants alien to Kyrgyzstan.

483   *Biodiversity Data Journal* 2:e1018.

484   Lejeune ALS, Courtois RJ. 1828. *Compendium florae Belgicae*. Leodii: P.J. Collardin.

485   Lososová Z. 2003. Estimating past distribution of vanishing weed vegetation in South Moravia.

486   *Preslia* 75:71–79.

487   Macfadyen S, Kriticos DJ. 2012. Modelling the Geographical Range of a Species with Variable

488   Life-History. *PLoS ONE* 7: e40313.

489   Meusel H, Jäger E, Weinert E. 1992. *Vergleichende Chorologie der Zentraleuropäischen Flora,*

490   *Vol. 1*. Jena: Gustav Fischer.

491 Mohr C. 1901. Plant Life of Alabama. In: *Contributions from the United States National*
492 *Herbarium*. Washington: US Department of Agriculture.

493 Newcomb S. 1994. A History of Chromite and Copper in Maryland: The Tyson Years. *Matrix*
494 3:84–92.

495 Phillips SJ, Dudík M, Schapire RE. 2004. A maximum entropy approach to species distribution
496 modeling. In: *Proceedings of the Twenty-First International Conference on Machine Learning.*
497 *ACM (New York)*. 655–662.

498 Petitpierre B, Kueffer C, Broennimann O, Randin C, Daehler C, Guisan A. 2012. Climatic niche
499 shifts are rare among terrestrial plant invaders. *Science* 335:1344–1348.

500 Plachy J. 1992. *The Mineral Industry of Norway in Minerals yearbook mineral industries of*
501 *Europe and central Eurasia*. U.S. Bureau of Mines.

502 Planchuelo AM. 1975. Study of the fruits and seeds of the genus Chenopodium in -Argentina.
503 *Darwiniana* 19:528–565.

504 Preston CD, Pearman DA, Dines TD .2002. New atlas of the British and Irish flora. An atlas of
505 the vascular plants of Britain, Ireland, the Isle of Man and the Channel Islands. New York:
506 Oxford University Press.

507 Reed CF. 1964. A flora of the chrome and manganese ore piles at Canton, in the Port of
508 Baltimore, Maryland and at Newport News, Virginia, with descriptions of genera and species new
509 to the flora of eastern United States. *Phytologia* 10:324–406.

510 Sautter G, Böhm K, Agosti D. 2007. Semi-Automated XML Markup of Biosystematics Legacy
511 Literature with the GoldenGATE Editor. In: *Proceedings of Pacific Symposium on Biocomputing*
512 *2007, Wailea (USA), January 2007*. International Society for Computational Biology: 391–402.

513 Schmidt-Lebuhn AN, Knerr NJ, Kessler M. 2013. Non-geographic collecting biases in herbarium
514 specimens of Australian daisies (Asteraceae). *Biodiversity and conservation* 22:905–919.

515 Šerá B. 2011. Stress tolerant plant species spread in the road-net. *Ecological Questions* 14:45–46.

516 Smart SM, Robertson JC, Shield EJ, van de Poll HAM. 2003. Locating eutrophication effects
517 across British vegetation between 1990 and 1998. *Global Change Biology* 9:1763–1774.

518 Soberón J, Peterson AT. 2011. Ecological niche shifts and environmental space anisotropy: a
519 cautionary note. *Revista Mexicana de Biodiversidad* 82:1348–1355.

520 Sutcliffe OL, Kay QON. 2000. Changes in the arable flora of central southern England since the

521 1960s. *Biological Conservation* 93:1–8.

522 Townsend Peterson A. 2011. Ecological niche conservatism: a time-structured review of

523 evidence. *Journal of Biogeography* 38:817–827.

524 Turner W. 1548. The names of herbes. (with an introduction, an index of English names, and an

525 identification of the plants enumerated by Turner) by James Britten. Published for the English

526 Dialect Society by N. Trübner 1881, London.

527 Uotila P. 2001. Chenopodium L. In: Jonsell B, ed. *Flora Nordica*. Stockholm: Bergius

528 Foundation, 4–31.

529 Van Landuyt W, Hoste I, Vanhecke L, Van den Bremt P, Vercruysse W, De Beer D. 2006. Atlas

530 van de Flora van Vlaanderen en het Brussels gewest. Nationale Plantentuin en het Instituut voor

531 Natuur- en Bosonderzoek i.s.m. Flo.Wer vzw.

532 Van Landuyt W, Vanhecke L, Hoste I, Hendrickx F, Bauwens D. 2008. Changes in the

533 distribution area of vascular plants in Flanders (northern Belgium): eutrophication as a major

534 driving force. *Biodiversity and Conservation* 17:3045–3060.

535 Vellend M, Brown CD, Kharouba HM, McCune JL, Myers-Smith IH. 2013. Historical ecology:

536 using unconventional data sources to test for effects of global environmental change. *American*

537 *Journal of Botany* 100:1294–1305.

538 Withering W. 1776. *A botanical arrangement of all the vegetables naturally growing in Great-*

539 *Britain Vol. 1*. London: Cadel and Elmsley.

540 Wróbel M, Tomaszewicz T. Chudecka J. 2006. Floristic Diversity and Spatial Distribution of

541 Roadside Halophytes along Forest and Field Roads in Szczecin Lowland (West Poland).

542 *Polish Journal of Ecology* 54:303–309.

543 Zengin Y. 1957. The Mode of Distribution of Chrome-Ores in Peridotites in Turkey. *Bulletin of*

544 *the Mineral Research and Exploration Institute of Turkey* 49:84–92.

545 Zhu G, Bu W, Gao Y, Liu G. 2012. Potential Geographic Distribution of Brown Marmorated

546 Stink Bug Invasion (Halyomorpha halys). *PLoS ONE* 7(2):e31246.

547 doi:10.1371/journal.pone.0031246

Introduction vectors gleaned from historical sources

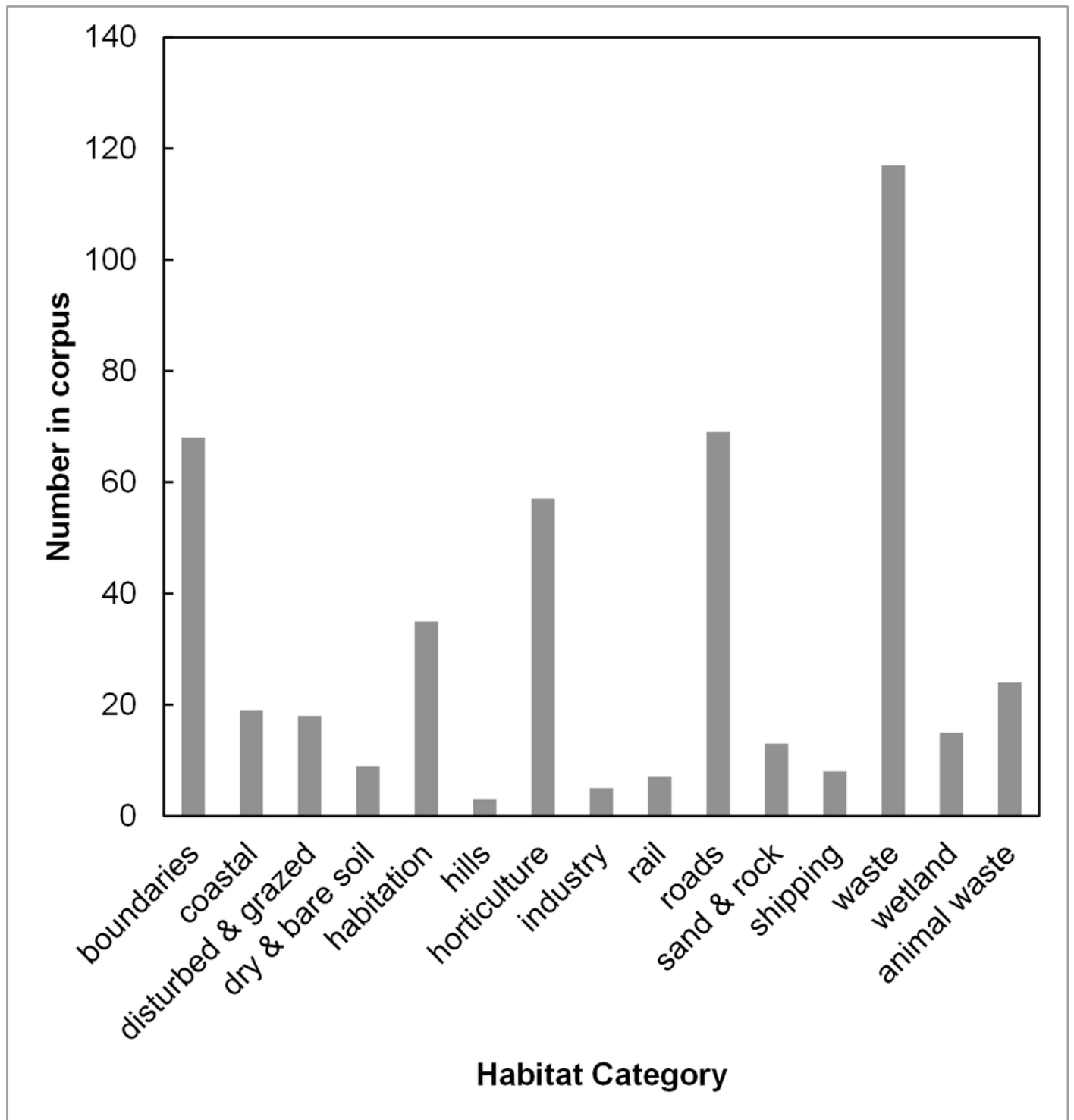
The vectors stated or implied from specimens and publications, including the range of dates that vectors were mentioned either on specimens or in publications.

Table 1. The introduction vectors stated or implied from specimens and publications, including the range of dates that vectors were mentioned on specimens and in publications.

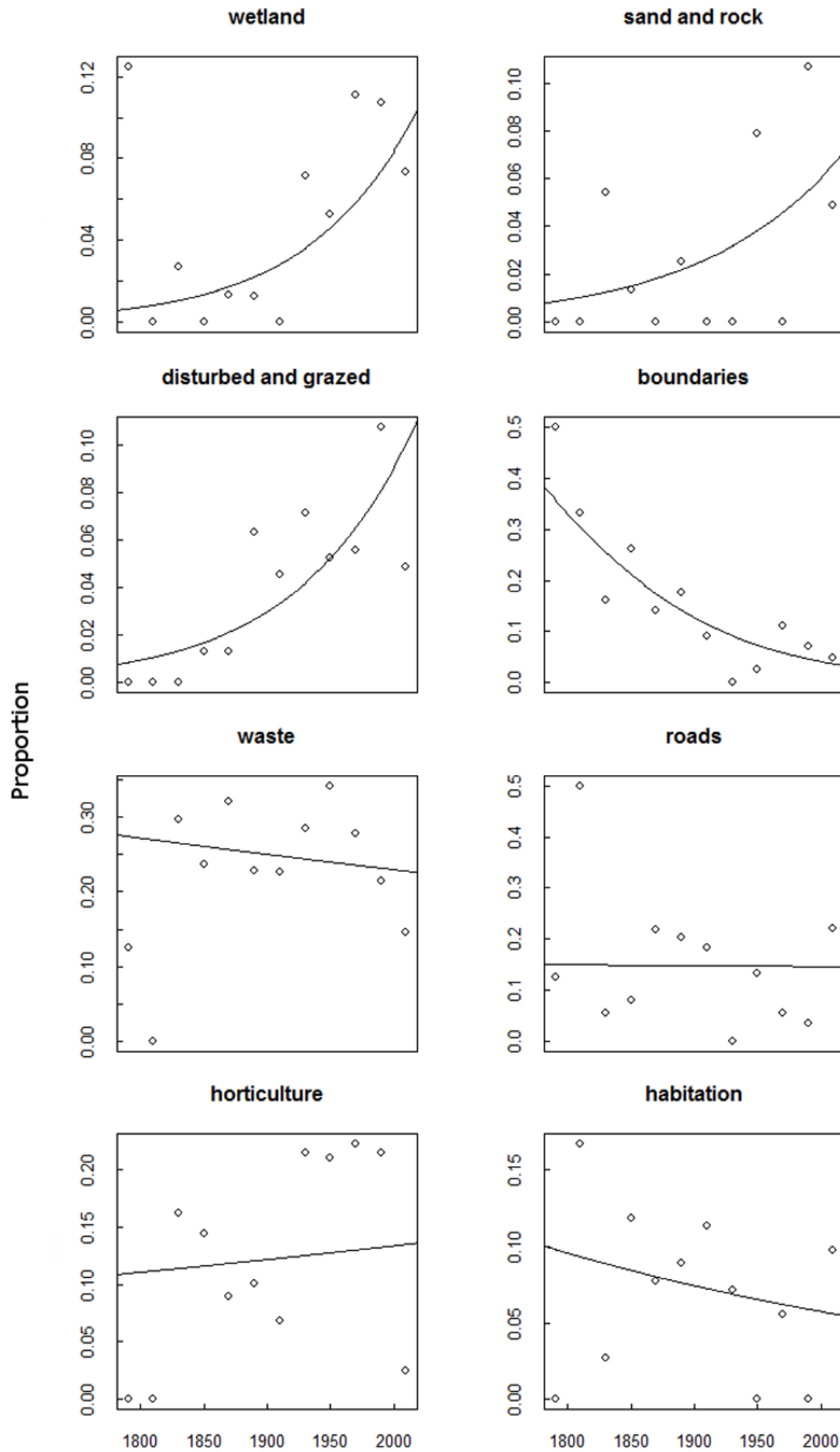
Vector	Dates	number	Example references and specimens
Ballast	1870 –1912	13	Publications: Burk, 1877; Mohr, 1901; Hjelt, 1906 Specimens: BIRM 032912; MANCH.94943.Kk803; S-H-2810
Grain	1936 –1964	3	Uotila, 2001; Unaccessioned specimens from Nationaal Herbarium Nederland (L);
Wool	1909	1	Observation by I.M. Hayward, Selkirkshire in database of the Botanical Society of Britain and Ireland (2013)
Ore	1931 –1961	4	Uotila, 2001; Reed, 1964; Lazkov, Sennikov & Naumenko, 2014; Specimen S-H-2141
Cork	1956 –1966	1	Uotila, 2001
Horticultural imports	2008	1	Hoste *et al*., 2009

1

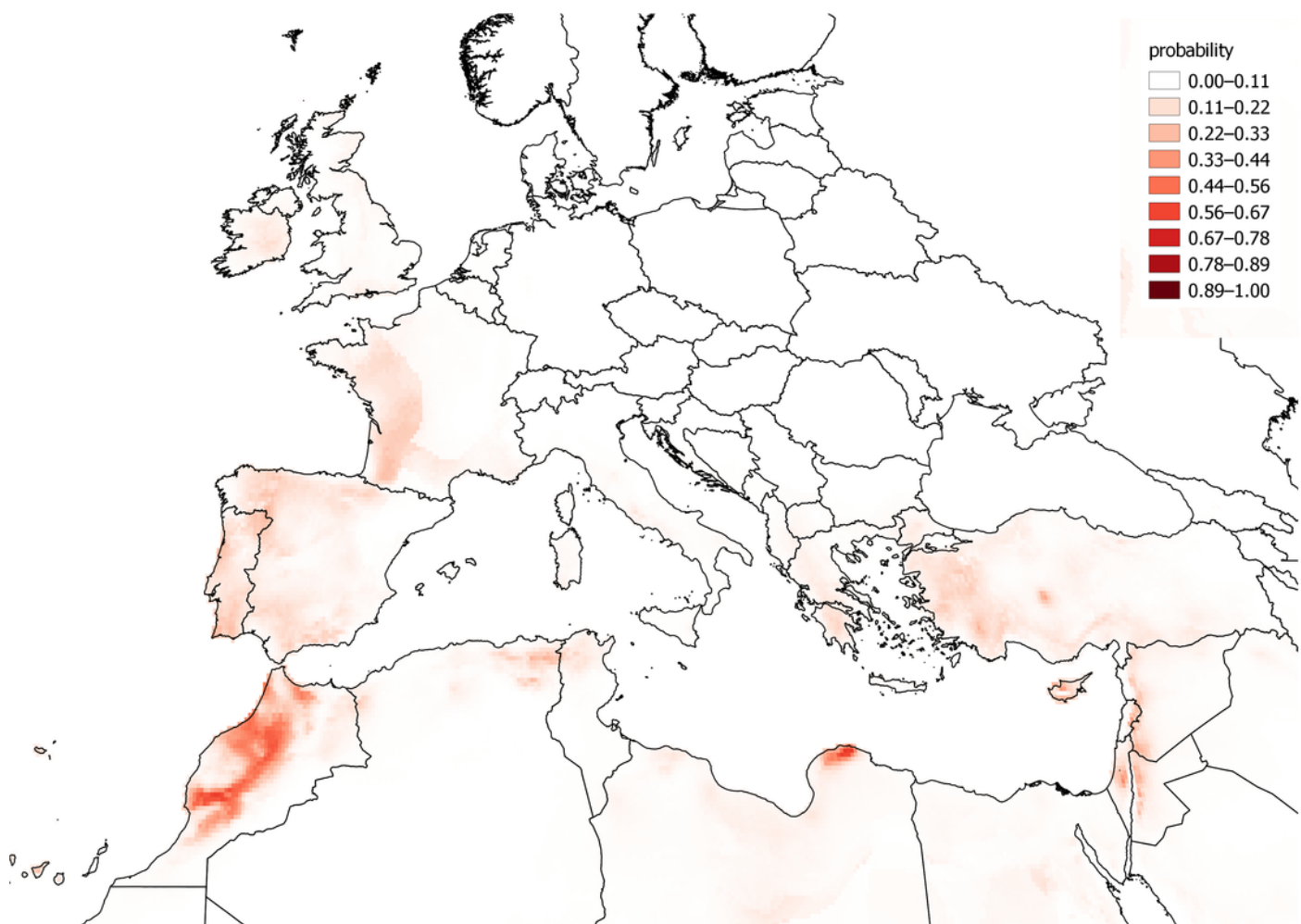
The use frequency of words in the collected corpus of *Chenopodium vulvaria* habitat descriptions

The frequency of each habitat category in the corpus of habitat descriptions from literature and specimens. The word and phrase tokens contributing to each category are presented in Table S3.

2

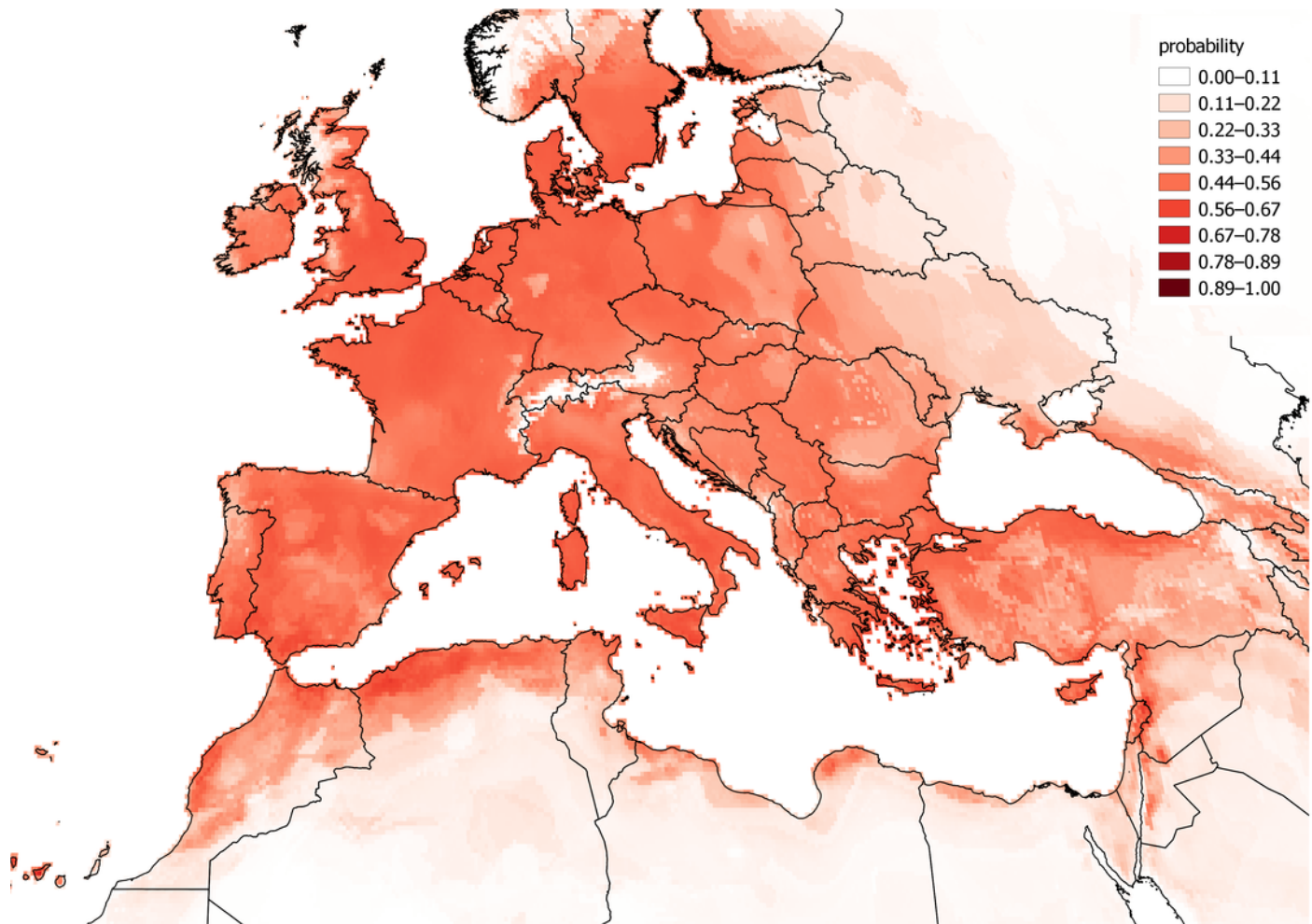
The change with time of habitat categories from the collected corpus of *Chenopodium vulvaria* habitat descriptions

A temporal analysis of the corpus of habitat descriptions from publications and specimens of *Chenopodium vulvaria*. The graphs show the proportion of token usage related to each habitat category for periods of 20 years. The words contributing to each habitat category are listed in Table S3. The best fit lines are from generalised linear models of the data weighted with the number of tokens contributing to each proportion. The categories wetland ($P < 0.01$, DF=11), sand and rock ($P < 0.05$, DF=11), and disturbed and grazed ($P < 0.05$, DF=11) all significantly increased with time. Only the term boundaries decreased with time ($P < 0.001$, DF=11). All other categories shown in Fig. 1 did not show significant variations with time.

3

A distribution model created from the naturalised range of *Chenopodium vulvaria* outside Europe and extrapolated back to Europe.

A distribution model of *Chenopodium vulvaria* in Europe, North Africa and the Middle-east projected from its naturalised range in California, South America and Australia. This model aims to predict where, according to the naturalised range, the climate is suitable for persistent populations in Europe as opposed to casual occurrences. The map uses a Mollweide equal area projection.

4

A distribution model of *Chenopodium vulvaria* created from all known locations

A distribution model of *Chenopodium vulvaria* in Europe, North Africa and the Middle-east created from all observations globally. This model aims to delimit the area where the climate is suitable for both stable populations and casual occurrences to occur. The map uses a Mollweide equal area projection.

A dated distribution map of *Chenopodium vulvaria* observations from Europe, North
Africa and the Middle-east

A distribution map of *Chenopodium vulvaria* in Europe, North Africa and the Middle-east.
Circles represent georeferenced observations either from specimens or from the literature.
The colour of the points denotes the date of observation, yet to emphasise the scarcer old
records the date ranges are not equal, but the data is divided into equal-sized subsets. The
map uses a Mollweide equal area projection.